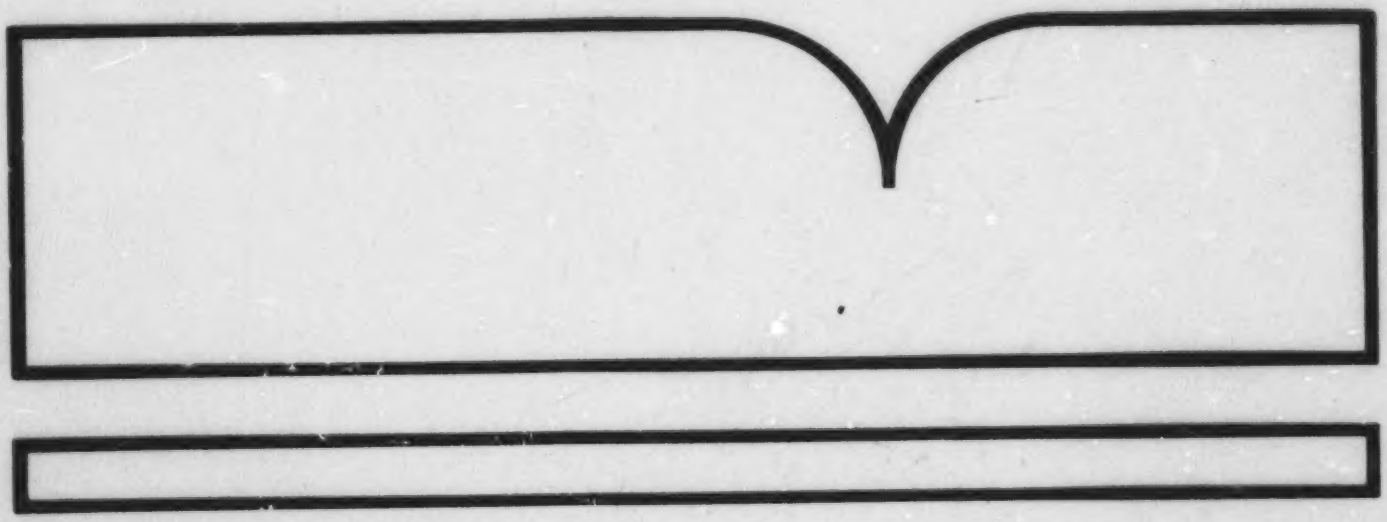
PB84-201623

Advantages of Using Regression Analysis to Calculate Results of Chronic Toxicity Tests

(U.S.) Environmental Research Lab.-Duluth, MN

Jun 84



U.S. Department of Commerce
National Technical Information Service
NTIS

PB84-201623

EPA-600/D-84-149
June 1984

Advantages of Using Regression Analysis to Calculate Results of Chronic Toxicity Tests

by

Charles E. Stephan

U.S. Environmental Protection Agency

Environmental Research Laboratory-Duluth

6201 Congdon Boulevard

Duluth, Minnesota 55804

and

John W. Rogers

Center for Lake Superior Environmental Studies

University of Wisconsin-Superior

Superior, Wisconsin 54880

ENVIRONMENTAL RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
DULUTH, MN 55804

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/D-84-149	2.
3. RECIPIENT'S ACCESSION NO. PB84 201623	
4. TITLE AND SUBTITLE Advantages of Using Regression Analysis to Calculate Results of Chronic Toxicity Tests	5. REPORT DATE June 1984 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) C. E. Stephan* and J. W. Rogers**	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS *Env. Res. Lab., USEPA, Duluth, MN 55804 **Center for Lake Superior Env. Studies University of Wisconsin-Superior Superior, WI 54880	10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Research Laboratory Office of Research and Development U.S. Environmental Protection Agency Duluth, MN 55804	13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE EPA-600/03

15. SUPPLEMENTARY NOTES
Proceedings of the 8th Symposium on Aquatic Toxicology, ASTM, 1984

16. ABSTRACT

Although it is traditional to calculate results of chronic toxicity tests using hypothesis testing to detect statistically significant differences from the control, calculation of results using regression analysis offers several major advantages. Regression analysis can directly produce a point estimate of the concentration causing any specified amount of adverse effect. In addition, whereas the degrees of freedom, precision of the test, and selection of have no impact on results calculated using regression analysis, they have a great impact on results calculated using hypothesis testing. Further, use of regression analysis emphasizes the necessity of considering the biological importance of the adverse effects observed in toxicity tests. By using hypothesis testing, aquatic toxicologists have been able to avoid this difficult, important question by presenting results in terms of a variable "statistically significant difference".

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT Release to Public	19. SECURITY CLASS *(This Report)* Unclassified 20. SECURITY CLASS *(This page)* Unclassified	21. NO. OF PAGES 19 22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Abstract:

Although it is traditional to calculate results of chronic toxicity tests using hypothesis testing to detect statistically significant differences from the control, calculation of results using regression analysis offers several major advantages. Regression analysis can directly produce a point estimate of the concentration causing any specified amount of adverse effect. In addition, whereas the degrees of freedom, precision of the test, and selection of α have no impact on results calculated using regression analysis, they have a great impact on results calculated using hypothesis testing. Further, use of regression analysis emphasizes the necessity of considering the biological importance of the adverse effects observed in toxicity tests. By using hypothesis testing, aquatic toxicologists have been able to avoid this difficult, important question by presenting results in terms of a variable "statistically significant difference".

Key Words: Aquatic toxicology, chronic tests, hypothesis tests, MATC, NOEC, regression analysis.

Results of almost all life-cycle, partial life-cycle, and early life-stage toxicity tests have been calculated using hypothesis testing to detect statistically significant differences from the control treatment. The apparent universal acceptance of the hypothesis testing approach is surprising, considering that no justification for its use has been published. The most plausible explanation is that aquatic toxicologists have followed the example of mammalian toxicologists who routinely use hypothesis testing to analyze data from chronic tests and many other sublethal tests. Another possible explanation is that many aquatic toxicologists do not know that regression analysis is a readily available, statistically sound alternative. The purpose of this paper is to explain several advantages of using regression analysis rather than hypothesis testing for calculating results of chronic toxicity tests with aquatic animals.

1. Regression analysis can directly produce point estimates of concentrations causing specified amounts of adverse effects.

In most chronic toxicity tests with aquatic animals, test organisms are subjected to about six treatments (a control and five concentrations of a test material) and data are collected on such things as the survival, growth, and reproduction of the animals. Usually a hypothesis testing procedure, such as analysis of variance followed by a multiple comparison procedure, is used to identify the highest concentration of test material that did not cause a reduction that was statistically significant from the control and the lowest concentration that did. Because it is easier to use point estimates for such things as calculating acute-chronic ratios and deriving water quality criteria, the pair of concentrations is usually converted to a point estimate either by using one or the other of the limits [1] or by taking an

arithmetic or geometric mean of the limits [2]. All of these are ad hoc procedures and none is very satisfying.

In regression analysis a curve is fit to the data and is used to obtain a point estimate directly. Logit and probit analysis procedures are common examples of regression analysis; these are often used to calculate results of acute toxicity tests and could be used to analyze dichotomous data from chronic tests. (Although regression analysis procedures are explicitly discussed herein, the same arguments also apply to any procedure which allows calculation of a point estimate and its variance.) Thus the first and most obvious reason for using regression analysis is that it can produce a point estimate directly.

2. Small changes in the statistical procedure can only cause small changes in results calculated using regression analysis but can cause large changes in results calculated using hypothesis testing.

Using regression analysis, changes in the calculated results will be roughly proportional to the magnitude of the change in the statistical procedure. A small change, such as changing from the probit method to the logit method, will only cause a slight change in the results. In contrast, different hypothesis tests may produce identical or very different results; when the results differ, the minimum difference will be the reciprocal of the dilution factor. For example, if the dilution factor is one-half, the results must differ by at least a factor of two. A factor of two is larger than the usual difference between results produced by different regression methods. In addition to the problem of choosing between different hypothesis testing procedures, results obtained using any one hypothesis test might depend on whether the data are transformed. Unfortunately there are few, if

any, hard and fast rules about whether data should be transformed. The importance of choosing between available options is reduced when regression analysis is used.

3. Results calculated using regression analysis are independent of the degrees of freedom and precision of the test.

If hypothesis testing is used to calculate results, the investigator who uses a larger number of experimental units and is more conscientious about details during the test will tend to make test materials look more toxic than the investigator who uses fewer experimental units and is less conscientious. If regression analysis is used to calculate results, the more conscientious investigator will tend to obtain tighter confidence limits, but on the average the two investigators will obtain the same point estimates. The difference between regression analysis and hypothesis testing is quite apparent if the impact of the accidental loss of one chamber during a test is considered. When hypothesis testing is used, loss of a chamber will reduce the degrees of freedom and the power of the test, and thus may change the results by a factor of two. When regression analysis is used, loss of one chamber will widen the confidence interval, but will not change the expected value of the point estimate.

4. The choice of α does not affect results obtained using regression analysis.

In regression analysis changing α (the significance level) will change the confidence limits on the results, but will not change the results themselves. When hypothesis testing is used, the results of a chronic test can often be changed by a factor of two merely by changing α from 0.05 to

0.02 or 0.10. Stated another way, a concentration in a chronic test resulting in P = 0.056 is considered to have caused no adverse effect, whereas another concentration resulting in P = 0.044 is considered to have caused an adverse effect. This dependence of hypothesis testing on α is very unfortunate because there is no good rationale for selecting α. Data from almost all chronic aquatic toxicity tests have been analyzed using hypothesis testing and $\alpha = 0.05$, but no rationale has been put forth to adequately justify why the results should not be made, for example, a factor of two higher or lower by using $\alpha = 0.02$ or 0.10.

Any serious consideration of what value of α to use should also consider what happens when data are obtained on several effects in a test and then each set of data is analyzed with $\alpha = 0.05$. If two or more individual effects are analyzed with $\alpha = 0.05$, then α for the overall toxicity test will be greater than 0.05. Thus the investigator must decide whether to set α for the individual effects or for the overall test. In most cases small changes in individual αs probably will not be important, but in some cases a small change in an α will change the result of the toxicity test by a factor of two. Often no easy solution is available for the problem of adjusting for multiple values of α, but this problem is not nearly as important as justifying the selection of a specific value of α in the first place. Is it reasonable to invest the necessary time and money in a chronic test and then analyze the data based on a vague decision about what percentage of the time it is acceptable to draw the wrong conclusion?

5. Regression analysis does not require treating experimental units as replicates if in fact they are not.

When hypothesis testing is used, concentrations of test material in replicates must be the same. If the concentrations are not the same, an average concentration must be used or else some or all of the data must be discarded. In regression analysis each experimental unit is treated independently and it is not necessary for the concentrations in the intended replicates to be the same. If differences occur, they can be accepted as real differences.

6. Use of regression analysis will discourage people from thinking that hypothesis testing identifies "no effect" concentrations.

Many people think of and discuss chronic toxicity in terms of a "no effect concentrations" [1, 3-5], although most readily acknowledge that the proper term is "no observed effect concentration", which actually means "no statistically significant effect concentration". Unfortunately some people who use data generated by aquatic toxicologists assume that "no effect concentration" actually refers to concentrations that do not cause an affect. Chronic toxicity may or may not be a threshold phenomenon, but even if it is, it is incorrect to assume that the "no statistically significant effect concentration" is the threshold. Because the power of chronic tests is variable and often quite low, 10-50% adverse effect may occur at concentrations that are reported to be no different from the controls [6-10]. If regression analysis is used to calculate results, it ought to be possible to develop a terminology that is compared to the LC50-EC50 terminology. Regardless of the terminology adopted, use of regression analysis should go a long way towards discouraging people from assuming that chronic tests identify "no effect concentrations".

7. Use of regression analysis will encourage people to think of chronic toxicity in terms of concentration-effect curves.

Because results of acute toxicity tests are usually calculated using regression analysis, whereas results of chronic toxicity tests are usually calculated using hypothesis testing, it may seem reasonable to assume that there is a fundamental difference between acute and chronic toxicity. Raw data from numerous chronic toxicity tests clearly indicate that it is appropriate to interpret chronic toxicity in terms of concentration-effect curves. Because results of acute toxicity tests have traditionally been interpreted in terms of concentration-effect curves, it is recommended that the concentrations used in an acute toxicity test produce a wide range of amounts of effect (see Section 12.2.2 of ASTM Standard E 729-80). No such recommendations have been proposed for chronic tests, probably because results are usually calculated using hypothesis testing.

The use of hypothesis testing to analyze data from chronic tests occasionally results in one concentration causing a statistically significant adverse effect even though a higher concentration does not produce a statistically significant adverse effect [3, 10-12]. Such data at a low concentration are usually ignored, but similar data at the highest tested concentration are usually accepted as a "real" effect even if no other supporting data are available. A way to avoid this problem is to treat chronic toxicity as a concentrationeffect phenomenon and require that the data indicate the existence of a concentration-effect relationship.

8. Use of regression analysis will encourage aquatic toxicologists to consider the real-world importance of the observed effects.

When regression analysis is used, a point estimate can only be obtained by specifying an absolute or relative amount of effect and calculating the best estimate of the concentration of test material that will produce that amount of effect. Regression analysis can end with an equation (with confidence limits) relating effect and concentration has been derived, and this may be appropriate in publications of results of chronic tests. Once the equation is available, anyone can calculate the concentration corresponding to any specified amount of effect. Almost all practical uses of the data from a chronic test will require the calculation of at least one such concentration, thus necessitating consideration of the real-world importance of the effect.

By using hypothesis testing, aquatic toxicologists have been able to avoid considering how much adverse effect is unacceptable, but only because they have ignored two of the three factors [13] that are important in hypothesis testing. Although aquatic toxicologists do select α (the level of type I error), they rarely if ever specify β (the level of type II error) or the minimum after a chronic test is completed, by specifying α and β, investigators could calculate the minimum amount of effect that would have been statistically significant. More importantly, before the test is begun the investigator should select α, β, and the minimum difference that is to be detectable as statistically significant, and then calculate the minimum necessary number of experimental units. Because chronic tests have not been specifically designed to be analyzed using hypothesis testing, rather large amounts of effect are not statistically significant in some tests [3, 7-9, 14], whereas rather small amounts of effect are significant in other tests [9, 10]. However, even if aquatic toxicologists were to select the minimum

amount of effect that should be detectable as statistically significant, hypothesis testing would not allow estimation of the concentration most likely to cause that selected amount of effect.

As discussed above, whether a particular amount of effect is statistically significant depends on a variety of things, but it does not depend at all on the biological importance of the amount of effect. If hypothesis testing can sometimes find an inconsequential effect to be statistically significant and sometimes cannot find severe adverse effects to be significant, why use hypothesis testing at all? Regardless of the outcome of hypothesis testing, it is still necessary to draw a separate conclusion concerning the biological importance of the observed effects.

By using regression analysis to obtain a point estimate, aquatic toxicologists can estimate the concentration most likely to result in an amount of adverse effect that was selected on the basis of biological importance. Determining the amount of adverse effect that should be considered unacceptable probably will not be easy but recognizing the importance of the problem will encourage people to work on it. It may be possible to reach some agreement without too much work. For example, most people would probably agree that a 50% reduction in the number of individuals reaching maturity would be unacceptable. And most people would similarly agree that even a 20% reduction in the average weight at maturity would be unacceptable. Another approach that might be used is ecosystem modeling [15, 16]. The important point is that recognizing the importance of this problem greatly increases the chances of making substantial progress toward solving it.

Previous Uses of Regression Analysis

Regression analysis is usually used to calculate results of acute toxicity tests, and regression analysis has been used to analyze data from some chronic tests. Brungs [17] fit a curve through data on the number of eggs per female. Although this is a regression-like approach, no statistics were used. Biesinger and Christensen [18] used the probit model to estimate the concentration corresponding to a 16% reduction in number of young produced. In this case the probit model was applied to data that are not dichotomous. In addition, regression analysis was used to simulate a form of hypothesis testing because "The 16% reproductive-impairment concentration is given since it represented a minimal reproducible value below which the variability in reproduction could not be detected from controls." Similarly, Pickering et al. [19] used regression analysis, but not the probit model, to estimate the highest concentration that would not have resulted in a statistically significant reduction in the number of eggs per female. Westlake et al. [20] used probit analysis to calculate the concentration which prevented 5% of the daphnids from producing free-swimming young on the assumption that "5% impairment of reproduction would be a negligible effect and therefore relatively 'safe'."

Feder and Collins [21, 22] proposed a way of using regression analysis to calculate results of chronic tests. However, rather than using the best estimate of the concentration that produces in the specified amount of adverse effect, they suggested using the lower 95% confidence limit on the concentration. The reason for using this lower concentration was to decrease the chance of the occurrence of the unacceptable amount of adverse effect. Unfortunately, this concept defeats several of the advantages of regression

analysis, because it makes the calculated result of the chronic test dependent on α, the degrees of freedom, and the precision of the investigator.

Although both the proposal of Feder and Collins and hypothesis testing make the result of the test dependent on aspects of experimental design and methodology, there is an important difference. With hypothesis testing, the poorer toxicity test, i.e., the test with fewer degrees of freedom and less agreement among replicate experimental units, will produce a higher result than the better test; however with the proposal of Feder and Collins, the poorer toxicity test will produce a lower result than the better test. Thus a lazy but knowledgeable investigator who wanted to make test materials look as toxic as possible would certainly prefer to calculate results of toxicity tests using the proposal of Feder and Collins rather than hypothesis testing. The direct use of regression analysis to obtain the best estimate of the concentration causing a specified absolute or relative amount of adverse effect eliminates these kinds of pitfalls. In the direct use of regression analysis the lazy investigator who uses a minimal number of experimental units or achieves poor reproducibility between replicate experimental units will conduct a test that is demonstrably poor because the confidence interval on the point estimate will be unusually large.

Apparent Problems With Using Regression Analysis

From a statistical standpoint, the major problem with using regression analysis to calculate results of chronic tests is selecting a model, i.e., the curve to fit through the points. The model selection process should involve input from both toxicologists and statisticians and analysis of available data. As with hypothesis testing, regression analysis should include checking that the assumptions and models used are appropriate to the

data. Fortunately, the importance of selecting an appropriate model can be reduced by requiring that point estimates be obtained by interpolation, not extrapolation. If the data requirements are reasonably stringent, point estimates calculated using different models will not differ very much. On the other hand, the stricter the data requirements, the more chronic tests that will not be good enough to allow calculation of a point estimate. But even in such cases, the chronic test should produce either an upper or lower limit on the point estimate. In the extreme if the chronic test does not produce enough good data, results should not be calculated using either hypothesis testing or regression analysis.

From a toxicological viewpoint, the major problem with using regression analysis is the necessity for selecting an absolute or relative amount of effect that is considered unacceptable. Indeed, this will be a formidable task. Ultimately though it will be worth the effort because the alternatives are so poor. Even hypothesis testing cannot be used in a sound manner without first selecting this critical amount of effect. Regression analysis does allow the option of calculating an equation from the results of a chronic test. Thus chronic tests can be conducted before a decision is made on the critical amount of effect. Once the equation is available different critical amounts of effect can be used in different situations without requiring that any chronic tests be rerun.

From an experimental viewpoint, the major problem with using regression analysis is that the imposition of stricter requirements on chronic tests will result in more chronic tests being deemed questionable or unacceptable. The counter argument, of course, is that the acceptance of poor tests merely reduces the credibility of aquatic toxicology. If hypothesis testing were to be used correctly, many chronic tests would be considered questionable or

unacceptable because they were not sensitive enough, i.e., did not have enough power to detect important differences.

From a historical viewpoint, the major problem with using regression analysis is that results of all previous chronic tests would have to be recalculated. This would obviously require much time and effort. On the other hand, some of the published tests are not good enough to warrant recalculating results using regression analysis, either because of the way the test was conducted or because of the data that were obtained. Acknowledging that advancements in aquatic toxicology have made some techniques obsolete should be interpreted as a sign of progress in aquatic toxicology. For example, in the tests reported by Mount and Stephan [23], a fairly high concentration of surfactant was used, the tests were begun with 1-inch fathead minnows, and the data were not analyzed statistically. A reexamination of many of the published chronic tests and recalculation of results, where appropriate, using regression analysis would require much time and effort but should be well worth the cost. Even if regression analysis is not adopted, all published chronic tests should be evaluated to see if the methods were acceptable and whether the data were analyzed appropriately.

From a regulatory viewpoint the major problem with using regression analysis is that most regulations and guidelines concerning chronic toxicity to aquatic organisms explicitly or implicitly are based on the use of hypothesis testing. On the other hand, good science should take priority over regulatory convenience. There is no valid reason why regulations and guidelines cannot be revised to be based on regression analysis, rather than hypothesis testing.

Summary

Regression analysis offers a variety of advantages over hypothesis testing for analyzing data from life-cycle, partial life-cycle, and early life-stage tests. The apparent disadvantages are all based on the fact that a substantial amount of effort will be necessary to implement the change. Defending and improving the use of hypothesis testing will also require much effort without producing substantial benefits. Switching from hypothesis testing to regression analysis will not be easy, but the benefits should be worth the effort.

References:

[1] Maki, A. W., Journal of the Fisheries Research Board of Canada, Vol. 36, 1979, pp. 411-421.

[2] U.S. Environmental Protection Agency, Federal Register, Vol. 45, 1980, p. 79345.

[3] McKim, J.M. and Benoit, D.A., Journal of the Fisheries Research Board of Canada, Vol. 31, 1974, pp. 449-452.

[4] Arthur, J.W., Lemke, A.E., Mattson, V.R. and Halligan, B.J., Water Research, Vol. 8, 1974, pp. 187-193.

[5] Spehar, R.L., Tanner, D.K. and Nordling, B.R., Aquatic Toxicology, Vol. 3, 1983, pp. 171-182.

[6] Hermanutz, R.O., Mueller, L.H. and Kempfert, K.D., Journal of the Fisheries Research Board of Canada, Vol. 30, 1973, pp. 1811-1817.

[7] Holcombe, G.W., Benoit, D.A., Leonard, E.N. and McKim, J.M., Journal of the Fisheries Research Board of Canada, Vol. 33, 1976, pp. 1731-1741.

[8] Hedtke, S.F. and Puglisi, F.A., Canadian Journal of Fisheries and Aquatic Sciences, Vol. 37, 1980, pp. 757-764.

[9] Biesinger, K.E., Anderson, L.E. and Eaton, J.G., Archives of Environmental Contamination and Toxicology, Vol. 11, 1982, pp. 769-774.

[10] Snarski, V.M. and Olson, G.F., Aquatic Toxicology, Vol. 2, 1982, pp. 143-156.

[11] Carlson, A.R., Journal of the Fisheries Research Board of Canada, Vol. 29, 1971, pp. 583-587.

[12] Spehar, R.L., Leonard, E.N. and DeFoe, D.L., Transactions of the American Fisheries Society, Vol. 107, 1978, pp. 354-360.

[13] Sokal, R.R. and Rohlf, F.J., Biometry, 1969, pp. 156-166.

[14] McKim, J.M. and Benoit, D.A., Journal of the Fisheries Research Board of Canada, Vol. 28, 1971, pp. 655-662.

[15] Waller, W.T., Dahlberg, M.L., Sparks, R.E. and Cairns, J., Jr., Journal of the Fisheries Research Board of Canada, Vol. 28, 1971, 1107-1112.

[16] O'Neill, R.V., Gardner, R.h., Barnthouse, L.W., Suter, G.W., Hildebrand, S.G. and Gehrs, C.W., Environmental Toxicology and Chemistry, Vol. 1, 1982, pp. 167-177.

[17] Brungs, W.A., Transactions of the American Fisheries Society, Vol. 98, 1969, pp. 272-279.

[18] Biesinger, K.E. and Christensen, G.M., Journal of the Fisheries Research Board of Canada, Vol. 29, 1972, pp. 1691-1700.

[19] Pickering, Q., Brungs, W. and Gast, M., Water Research, Vol. 11, 1977, pp. 1079-1083.

[20] Westlake, G.F., Sprague, J.B. and Rowe, D.W., Aquatic Toxicology, Vol. 4, 1983, pp. 327-339.

[21] Feder, P.I. and Collins, W.J., in Aquatic Toxicology and Hazard Assessment: Fifth Conference, ASTM STP 776, American Society for Testing and Materials, Philadelphia, 1982, pp. 32-68.

[22] Feder, P.I. and Collins, W.J., "Design and Analysis of Chronic Aquatic Tests of Toxicity." ADA 134309. National Technical Information Service, Springfield, VA. 1981.

[23] Mount, D.I. and Stephan, C.E., Transactions of the American Fisheries Society, Vol. 96, 1967, pp. 185-193.